AD-A259 155

(2)

POISSON COMPOUNDING OF DEPENDENT RANDOM VARIABLES: A STOCHASTIC MODEL FOR TOTAL CLAIM COSTS

J. E. Angus

Report No. 92-14

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
SAN DIEGO, CALIFORNIA 92186-5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

POISSON COMPOUNDING OF DEPENDENT RANDOM VARIABLES:

A STOCHASTIC MODEL FOR TOTAL CLAIM COSTS

John E. Angus

Department of Mathematics
The Claremont Graduate School
143 E. Tenth Street
Claremont, CA 91711-3988

and

Medical Information Systems and Operations Research Department
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

August 1992

Report No. 92-14, supported by the Naval Health Research Center under the American Society for Engineering Education Navy Summer Faculty Research Program. The views expressed in this article are those of the author and do not reflect the official policy or position of the Department of Defense, nor the U.S. Government. Approved for public release, distribution unlimited.

SUMMARY

The compound Poisson process is a useful model for describing total claim costs in the insurance industry. In this investigation, the compound Poisson model is modified to allow dependence among the compounding variables in an effort to more accurately model situations in which successive claim awards are correlated and/or form a nonstationary process. Specifically, the sequence is assumed to follow an ARIMA(p,d,q) model. This type of model is illustrated with the use of actual asbestosis claim cost data collected from Naval shipyards.

POISSON COMPOUNDING OF DEPENDENT RANDOM VARIABLES:

A STOCHASTIC MODEL FOR TOTAL CLAIM COSTS

JOHN E. ANGUS

Department of Mathematics
The Claremont Graduate School
143 E. Tenth Street
Claremont, CA 91711

1. Introduction and Preliminary Development

On a fixed probability space $(\Omega, \mathfrak{I}, P)$, let $\{N_T; T \geq 0\}$ be a nonhomogeneous Poisson process with mean value function $EN_T \equiv m(T)$, $T \geq 0$, satisfying $m(0) = 0$, $m(\bullet)$ is nondecreasing, $m(T) \uparrow \infty$ as $T \to \infty$, and let $\{X_t, t \geq 1\}$ be a sequence of random variables, independent of the Poisson process. When the X_is are independent and identically distributed (iid), the process $\{C_T, T \geq 0\}$ defined by

$$C_T = \sum_{t=1}^{N_T} X_t \tag{1}$$

(where a summation with upper index 0 is defined to be 0) is called a Compound Poisson Process (CPP). The CPP is used extensively in modeling the total claim costs in insurance risk analysis. See Prabhu (1980), for example. In this context, claims are generated according to a Poisson process, and successive claim awards are represented by the X_is. When applied to specific types of claims, especially those involving bodily injury, the CPP may no longer be an accurate portrayal of the process because the sequence of X_is may be autocorrelated and / or nonstationary. This could be caused, for example, by the effects of legal precedent established in adjudicating the claims, economic trends, and so on.

To account for possible trends and autocorrelation within the sequence of X_is, the CPP can be modified by assuming that $\{X_t, t \geq 1\text{-}d\}$ is an ARIMA(p, d, q) process. That is, introducing the back-shift operator B ($B^j X_t = X_{t-j}$) and difference operator ∇ ($\nabla^0 X_t = X_t$, $\nabla X_t = (1-B)X_t = X_t - X_{t-1}$, $\nabla^j X_t = \nabla(\nabla^{j-1} X_t)$, $j \geq 1$), it is assumed that for $t \geq 1$,

$$\nabla^d X_t = \mu + Y_t \tag{2}$$

where $\{Y_t\} \equiv \{Y_t, t = 0, \pm 1, \ldots\}$ is a zero-mean causal ARMA(p, q), and μ is a constant. This in turn means that $\{Y_t\}$ is a zero-mean second order stationary sequence that satisfies a set of difference equations of the form

$$\varphi(B)Y_t = \theta(B)Z_t \tag{3}$$

where $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p$, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$, and the polynomial $\varphi(\bullet)$, when treated as a function of a complex variable z, has no roots inside or on the unit circle in the complex plane. The sequence $\{Z_t\}$ is assumed to be a white noise sequence, that is, $EZ_t=0$, and $EZ_tZ_s = \sigma^2\delta(t-s)$ for all integers s and t, where $\delta(k) = 0$ if $k \neq 0$, and $\delta(0) = 1$. This is usually denoted by $\{Y_t\}$~ARMA(p, q), and $\{Z_t\}$~WN(0, σ^2).

There are, of course, infinite choices for the process $\{X_t\}$ that would account for autocorrelation and / or nonstationarity. The ARIMA choice is particularly attractive for many reasons. For example, it is known that ARMA (p, q) process, the d = 0 case, can approximate a broad class of stationary processes whose autocorrelation functions approach 0 as the lag approaches ∞. In addition, the ARIMA process describes processes that possess a deterministic or stochastic polynomial trend, as well as explosive - type nonstationarity (as exemplified by the simple random walk model). Finally, statistical techniques for identifying, estimating, diagnosing, and forecasting ARIMA models are well-developed, mature (see Box and Jenkins, (1970), Priestley (1981), or Brockwell and Davis (1992), for example), and highly automated in popular statistical computing packages, making them imminently practical for this application and others in which dependent sequences must be modeled.

To demonstrate the plausibility of this approach to modeling actual claim costs, 221 successive asbestosis claims collected from US Naval shipyards were compiled. These are listed in Table 2 of the Appendix. These are plotted in Figure 2. Figures 3 and 4 show, respectively, the estimated autocorrelation and partial autocorrelation functions of the series after differencing once at lag 1. These suggest that perhaps an MA(1), AR(5) or mixed model would be appropriate to describe the differenced series. Figure 5 shows the estimated autocorrelation function of the estimated residuals from a fitted ARIMA(0, 1, 1) model for the original series, which yield a Portmanteau test on the first 20 values of 15.01 on 19 degrees of freedom, which appears consistent with whiteness of the residuals and adequacy of the fit. The fitted model is

$$X_t = X_{t-1} + 244.13 + Z_t - .93\, Z_{t-1},\ \{Z_t\} \sim \mathrm{WN}(0, 6.04 \times 10^8).$$

Figure 1 is a plot of cumulative number of asbestosis claims versus time reported from the same Naval Shipyards. These are not in one-to-one correspondence with the claim amounts in Table 2 because not all claims had been adjudicated. Figure 1 also contains a fitted 5th degree polynomial approximating the mean value function of a postulated underlying nonhomogeneous Poisson process. Finally, Table 1 gives observed counts versus expected, computed from the postulated nonhomogeneous Poisson model, for randomly selected time intervals. Again, the data appear consistent with a nonhomogeneous Poisson process.

Without further assumptions on the sequence $\{Z_t\}$, distribution theory for $\{C_T\}$ cannot be addressed. However, assuming only that $\{Z_t\} \sim \mathrm{IID}(0, \sigma^2)$, asymptotic distributions for C_T, for large T, can be developed. This is the subject of the next section. The final section discusses how the asymptotic results may be used in practice.

2. Asymptotic Distribution theory for C_T

By the assumptions on $m(T) = EN_T$, asymptotic distribution theory for C_T depends on the following results from Angus (1992). Let $\{X_t, t=1-d, 2-d, .., 0, 1, 2, ...\}$ be an ARIMA(p,d,q) process, satisfying $\nabla^d X_t = Y_t + \mu$ where $\{Y_t\}$ is a causal ARMA(p,q) process with $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$. Define the usual operation of generating factorial polynomials by $k^{(j)} \equiv \prod_{i=1}^{j} (k-i+1) = k(k-1)...(k-j+1)$ for integers j and k with $j>0$, and $k^{(0)} \equiv 1$. Thus, with k treated as a variable, it follows that for $j \geq 0$,

$$\nabla\, (k+1)^{(j+1)}/[(j+1)!] = k^{(j)}/j!, \tag{4}$$

and

$$\sum_{k=0}^{n} k^{(j)}/j! = \frac{(n+1)^{(j+1)}}{(j+1)!}. \tag{5}$$

The main results from Angus (1992) are the following theorem and corollary, which hold even for the case d=0 with the convention that summations with upper limit 0 are taken to be 0. The symbol "$\to$" denotes convergence in distribution.

Theorem 1. Let $\gamma_Y(\bullet)$ be the autocorrelation function of $Y_t=\nabla^d X_t-\mu$, and suppose that $\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)>0$. Then, as $n\to\infty$, the distribution of

$$\frac{(1/d!)\sum_{t=1}^{n}\left(X_t-\sum_{i=0}^{d-1}(\nabla^i X_0)(t+i-1)^{(i)}/i! - \mu(t+d-1)^{(d)}/d!\right)}{\left[\gamma_Y(0)\sum_{v=1}^{n}\left[(v+d-1)^{(d)}\right]^2 + 2\sum_{h=1}^{n-1}\gamma_Y(h)\sum_{v=1}^{n-h}(v+d-1)^{(d)}(v+h+d-1)^{(d)}\right]^{1/2}} \tag{6}$$

converges to that of N(0,1).

Corollary 1. Under the conditions of Theorem 1, $n^{-1/2-d}\sum_{t=1}^{n}\left(X_t-\mu(t+d-1)^{(d)}/d!\right)\to N(0,\omega^2)$ as $n\to\infty$, where

$$\omega^2=(2d+1)^{-1}d!^{-2}\left[\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)\right]. \tag{7}$$

The proofs of these are given in Angus (1992), and make use of the following Lemma, also proved in Angus (1992), that will be of further use in the present discussion.

Lemma 1. Suppose that $\nabla^d X_t=Y_t+\mu$, $t\geq 1$. Then for $t\geq 1$, X_t can be expressed as

$$X_t=\sum_{i=0}^{d-1}(\nabla^i X_0)\frac{(t+i-1)^{(i)}}{i!}+\frac{\mu(t+d-1)^{(d)}}{d!}+\sum_{v=1}^{t}Y_{t-v+1}\frac{(v+d-2)^{(d-1)}}{(d-1)!} \tag{8}$$

and for $n\geq 1$

$$\sum_{t=1}^{n}X_t=\sum_{i=0}^{d-1}(\nabla^i X_0)\frac{(n+i)^{(i+1)}}{(i+1)!}+\frac{\mu(n+d)^{(d+1)}}{(d+1)!}+\sum_{v=1}^{n}Y_{n-v+1}\frac{(v+d-1)^{(d)}}{d!} \tag{9}$$

The next theorem gives the asymptotic distribution of C_T as $T\to\infty$. Here, the following notation will be used. For sequences a_n and $b_n>0$, T_n is $AN(a_n, b_n)$ means that

$$\lim_n P\left\{\frac{T_n-a_n}{\sqrt{b_n}}\leq t\right\}=(2\pi)^{-1/2}\int_{-\infty}^{t}\exp(-x^2/2)dx$$

for all real t, where n may be a discrete or continuous variable, and the limit is taken as $n\to c$ where c is real or $\pm\infty$.

Theorem 2. Let $\{X_t\}$ in (1) be an ARIMA(p, d, q) as in Theorem 1, and assume the conditions and notation given there. Then, as $T\to\infty$, C_T is $AN(\nu_T, \tau_T^2)$ where

$$\nu_T = \frac{\mu[m(T) + d]^{(d+1)}}{(d+1)!} \tag{10}$$

and

$$\tau_T^2 = \frac{m(T)^{2d+1}}{d!^2(2d+1)}\left(\gamma_Y(0) + 2\sum_{h=1}^{\infty}\gamma_Y(h) + \mu^2(2d+1)\right). \tag{11}$$

From Lemma 1, by conditioning on $\{X_{1-d}, X_{2-d}, ..., X_0\}$, the following corollary is immediate.

Corollary 2. Under the conditions of Theorem 1, if $d \geq 1$, then conditioned on $\{X_{1-d}, X, ..., X_0\}$, C_T is $AN(\nu_T, \tau_T^2)$ where

$$\nu_T = \sum_{i=0}^{d-1}(\nabla^i X_0)\frac{(m(T)+i)^{(i+1)}}{(i+1)!} + \frac{\mu[m(T)+d]^{(d+1)}}{(d+1)!} \tag{12}$$

and τ_T^2 is given in (11).

Another useful corollary is the special case d=0, which indicates that the process $\{X_t\}$ is a causal ARMA(p, q) process with mean μ, having autocovariance function $\gamma_Y(h) = \text{cov}(X_t, X_{t+h})$.

Corollary 3. Under the conditions of Theorem 2 with d=0, as $T \to \infty$, C_T is $AN(\nu_T, \tau_T^2)$ where

$$\nu_T = \mu\, m(T) \tag{13}$$

and

$$\tau_T^2 = m(T)\left(\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h) +\mu^2\right) \tag{14}.$$

Proof of Theorem 2. From (9), define the polynomial p(n) with random coefficients by $p(n) \equiv \sum_{i=0}^{d-1}(\nabla^i X_0)\frac{(n+i)^{(i+1)}}{(i+1)!} + \frac{\mu(n+d)^{(d+1)}}{(d+1)!}$. Then with ν_T and τ_T^2 as in (10) and (11) it follows that

$$\frac{C_T - \nu_T}{\tau_T} = \frac{\sum_{t=1}^{N_T} X_t - p(N_T)}{\tau_T} + \frac{p(N_T) - \nu_T}{\tau_T}. \tag{15}$$

By independence of thı processes $\{X_t\}$ and $\{N_T\}$, the two random variables on the right side of (15) are uncorrelated. By Theorem 1 and Corollary 1, the first term on the right side of (15) converges in distribution to

$$N_1(0,1)\left[\frac{\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)}{\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)+\mu^2(2d+1)}\right]^{1/2},$$

and by a first order Taylor expansion of the function $f(x) = (x+d)^{(d+1)}$, the second term on the right side of (15) converges in distribution to

$$N_2(0,1)\left[\frac{\mu^2(2d+1)}{\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)+\mu^2(2d+1)}\right]^{1/2},$$

where $N_1(0,1)$ and $N_2(0,1)$ denote independent standard normal random variables. The result now follows immediately.

3. Application

In practice, when T is large, Theorem 2 can be used to approximate the distribution of C_T. If all parameters are known, then the Theorem can be applied directly. Otherwise, the parameters must be estimated from data. The parameters that must be estimated include the mean value function m(T) for the Poisson process, and the parameters in the ARIMA(p, d, q) process, namely, p, d, q, $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q$, and σ^2. Efficient estimation of these inn turn leads to efficient estimation of the autocovariance function $\gamma_Y(\bullet)$. Appropriate data for accomplishing this are those listed in the Appendix for the asbestosis claims. For illustrative purposes, assume that the claims are generated according to an homogeneous Poisson process. Then for this dataset, m(•) would be estimated by

$$\hat{m}(T) = 1.63\,T. \tag{16}$$

As previously discussed, the series of claim amounts fits an ARIMA(0,1,1) model

$$X_t = X_{t-1} + \mu + Z_t + \theta_1 Z_{t-1},$$

with μ=244.13, θ_1= -.93, and σ^2=6.04 x 10^8. The autocovariance function of $Y_t = \nabla X_t$ in this case is

$\gamma_Y(0) = \sigma^2(1 + \theta_1^2)$, $\gamma_Y(\pm 1) = \sigma^2\theta_1$, $\gamma_Y(h) = 0$, for $|h|>1$.

Thus, (10) and (11) become

$$\nu_T = \frac{244.13(1.63T + 1)(1.63T)}{2} = 324.31\, T^2 + 198.97\, T \tag{17}$$

$$\tau_T^2 = \frac{(1.63T)^3}{3}\Big((6.04 \times 10^8)\big(1 + (-.93)^2 + (2)(-.93)\big) + (244.13)^2(3)\Big) \tag{18}$$

$$= (4.5 \times 10^6)\, T^3$$

so that in this example, C_T is AN(ν_T, τ_T^2) with ν_T and τ_T^2 given as in (17) and (18). This can be compared to the approximation that would be obtained if the claim data were treated naively as IID. If this were done, the sample mean of the X_i would be 15098.16, the sample variance would be 5.62×10^8, and the approximation would be C_T is AN(ψ_T, ξ_T^2) where

$$\psi_T = 15098.16(1.63T) = 24610.00\, T \tag{19}$$

$$\xi_T^2 = (1.63T)\big(5.62 \times 10^8 + (15098.16)^2\big) = (1.3 \times 10^9)\, T. \tag{20}$$

The discrepancy between (17) and (19), and between (18) and (20) is large, with (19) and (20) being of the wrong order of magnitude in T.

4. Conclusions

A generalization of the CPP model for insurance claim costs has been presented and its plausibility demonstrated using asbestosis claim data collected from Naval shipyards. Asymptotic distribution theory has been developed to allow approximation of the distribution of total claim costs during [0, T] for large T. These approximations and a numerical example demonstrate that large discrepancies can be experienced when applying the naive CPP when the more general model using an ARIMA(p,d,q) process is appropriate.

References

Angus, J.E. (1992). A central limit theorem for autoregressive integrated moving average processes. Report No. 92-15, Naval Health Research Center, San Diego, California.

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden Day: San Francisco.

Brockwell, P.J. and Davis, R.A. (1992). Time Series: Theory and Methods. New York:

Springer-Verlag.

Durrett, R. (1991). Probability: Theory and Examples. Pacific Grove: Wadsworth & Brooks/Cole.

Prabhu, N.U (1980). Stochastic Storage Processes: Queues, Insurance Risk, and Dams. New York: Springer Verlag.

Priestley, M.B. (1981). Spectral Analysis and Time Series, Vols. I and II. New York: Academic Press.

Appendix: Tables and Figures

Table 1: Observed Claims Versus Expected

Time	Interval	Observed (O)	Expected (E)	$(O-E)^2/E$
2	62	117	120.814	0.12041
62	71	16	19.953	0.78331
71	132	136	135.116	0.00579
132	145	27	27.534	0.01037
145	195	105	98.055	0.49184
195	222	45	47.070	0.09104
222	244	28	35.278	1.50164
244	298	77	76.380	0.00504
298	351	79	65.320	2.86491
351	450	107	119.842	1.37617
450	490	54	54.553	0.00560
490	520	48	44.331	0.30363
520	544	35	37.362	0.14936

Chi-square = 7.71 on 6 degrees of freedom (6 parameters were estimated).

Table 2

Asbestosis Claim Costs: Naval Shipyards, January 1990 through June 1991

(Read Data Across)

853	1082	1736	12035	29201	35473	42308
27454	29933	15832	29710	17809	2186	3803
39355	23214	609	10835	44982	16445	1946
2192	18571	93514	24097	561	24786	679
23975	17117	8782	6365	58995	948	17734
656	21587	49213	6369	1587	68921	24389
335	14456	124	1445	137	578	30277
1014	9848	660	7740	30045	2563	9511
542	21303	18366	11408	26127	25944	13553
17096	165	455	34596	672	389	7930
40184	3067	44565	74062	1216	368	26727
51622	673	30781	1325	22	10737	21835
828	19393	24764	1004	30319	1111	50087
555	593	33591	8659	1075	26805	1632
25730	2426	15152	505	375	11065	524
1768	23327	65	25336	557	1374	27221
8350	1932	191	2003	237	547	17190
265	45838	3275	638	2461	742	29577
15718	1045	15925	60	250719	2569	18267
432	448	26962	53647	19201	1996	60
1858	27388	1331	528	356	17492	15366
24614	30735	457	720	265	838	20073
11557	4388	566	44606	26693	15796	415
266	21380	24891	27628	22	33982	21919
37148	26608	15843	16177	49743	3802	17555
1381	1052	14444	903	753	555	16512
5227	695	44009	9353	420	198	993
108199	248	1436	1894	991	1710	457
5377	29479	60	32700	425	4511	41080
519	20725	16417	969	22484	1959	35298
245	1257	634	2016	561	402	550
950	385	255	54561			

Figure 1.
Cumulative Asbestosis Claims Versus Time
Solid Line is Least-Squares 5th Degree Polynomial

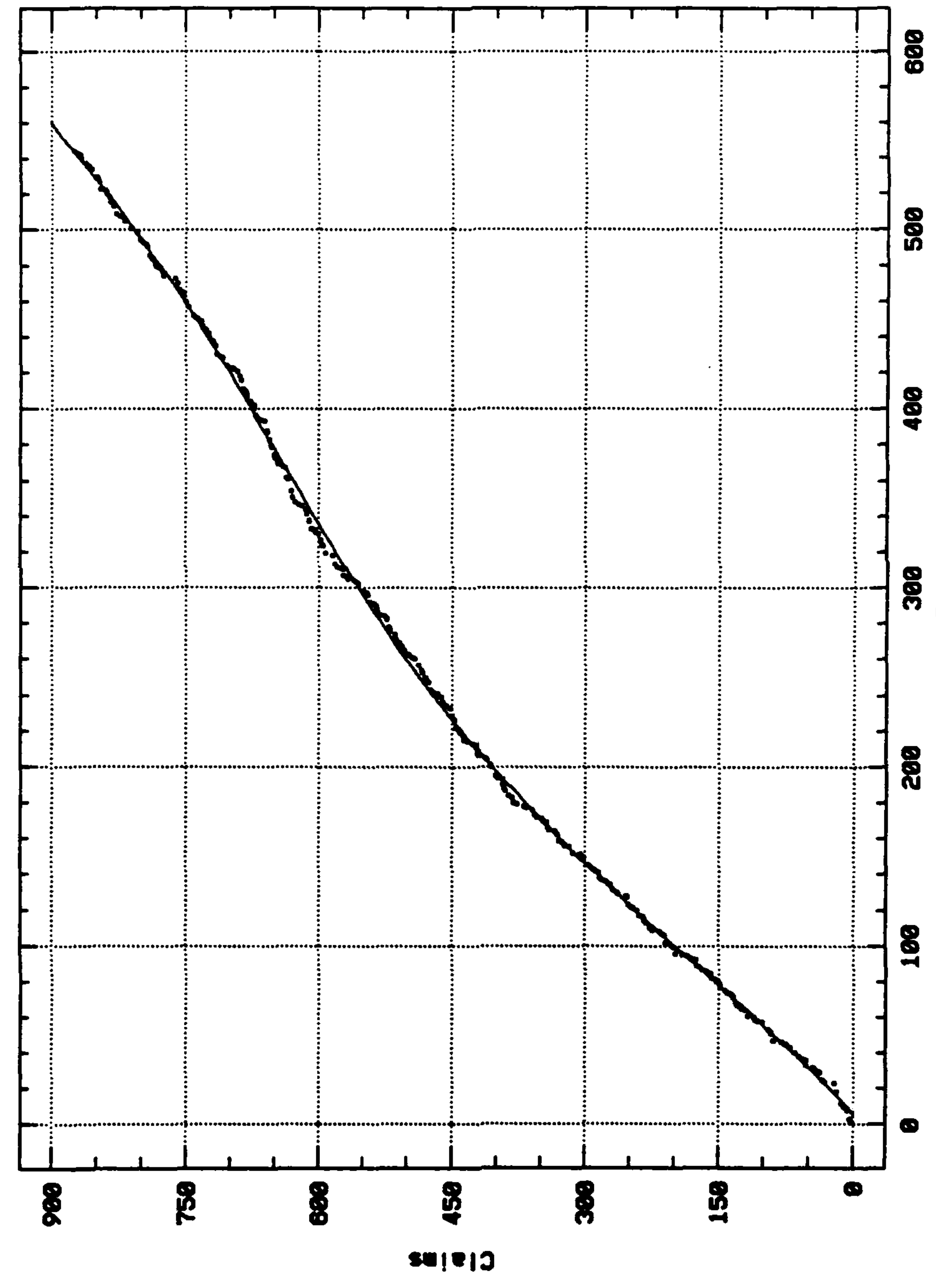


Figure 2.
Total Claim Costs
Asbestosis-Related Claims

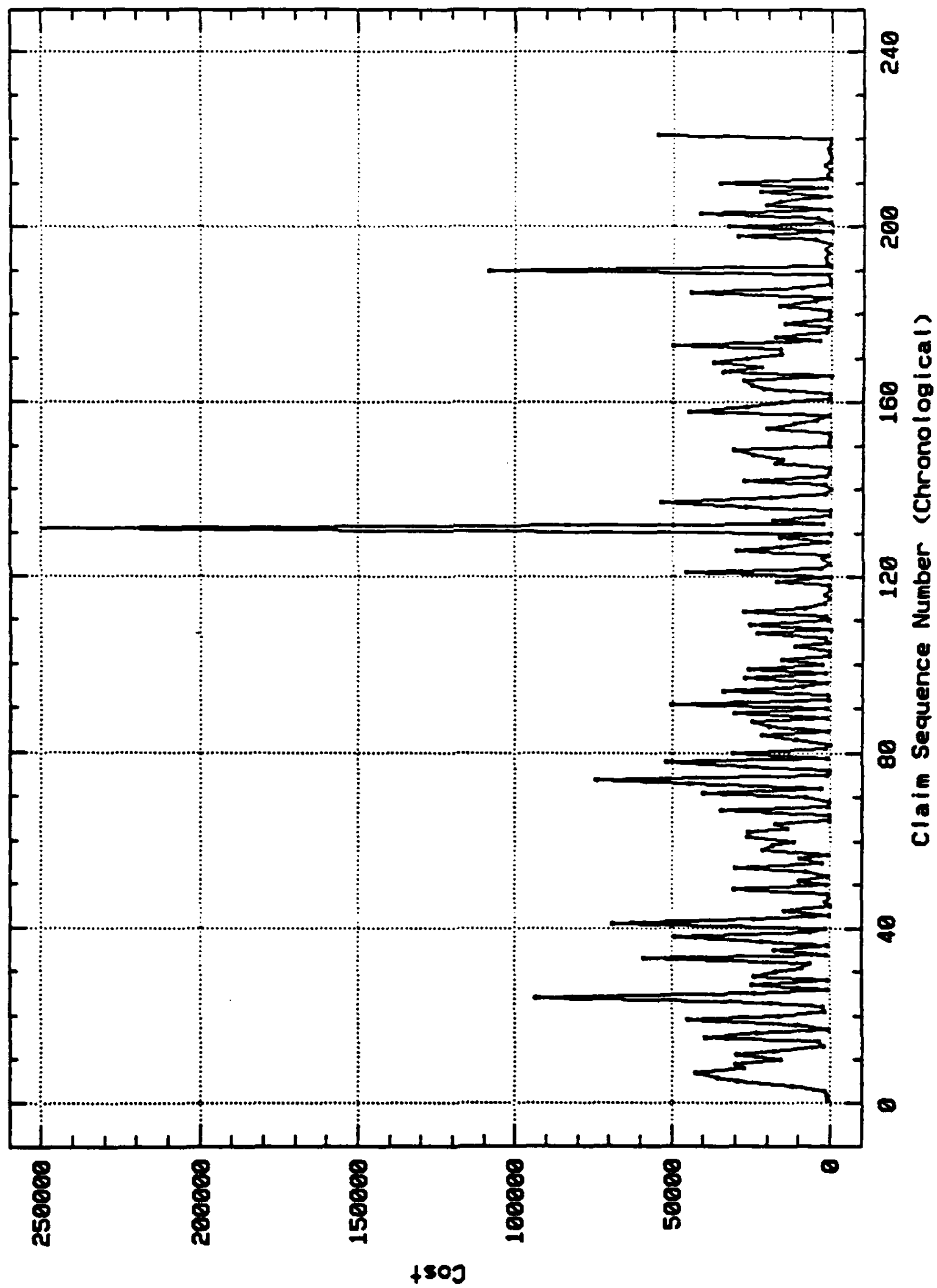


Figure 3.
Autocorrelations
Asbestosis Claim Costs (Differenced at Lag 1)

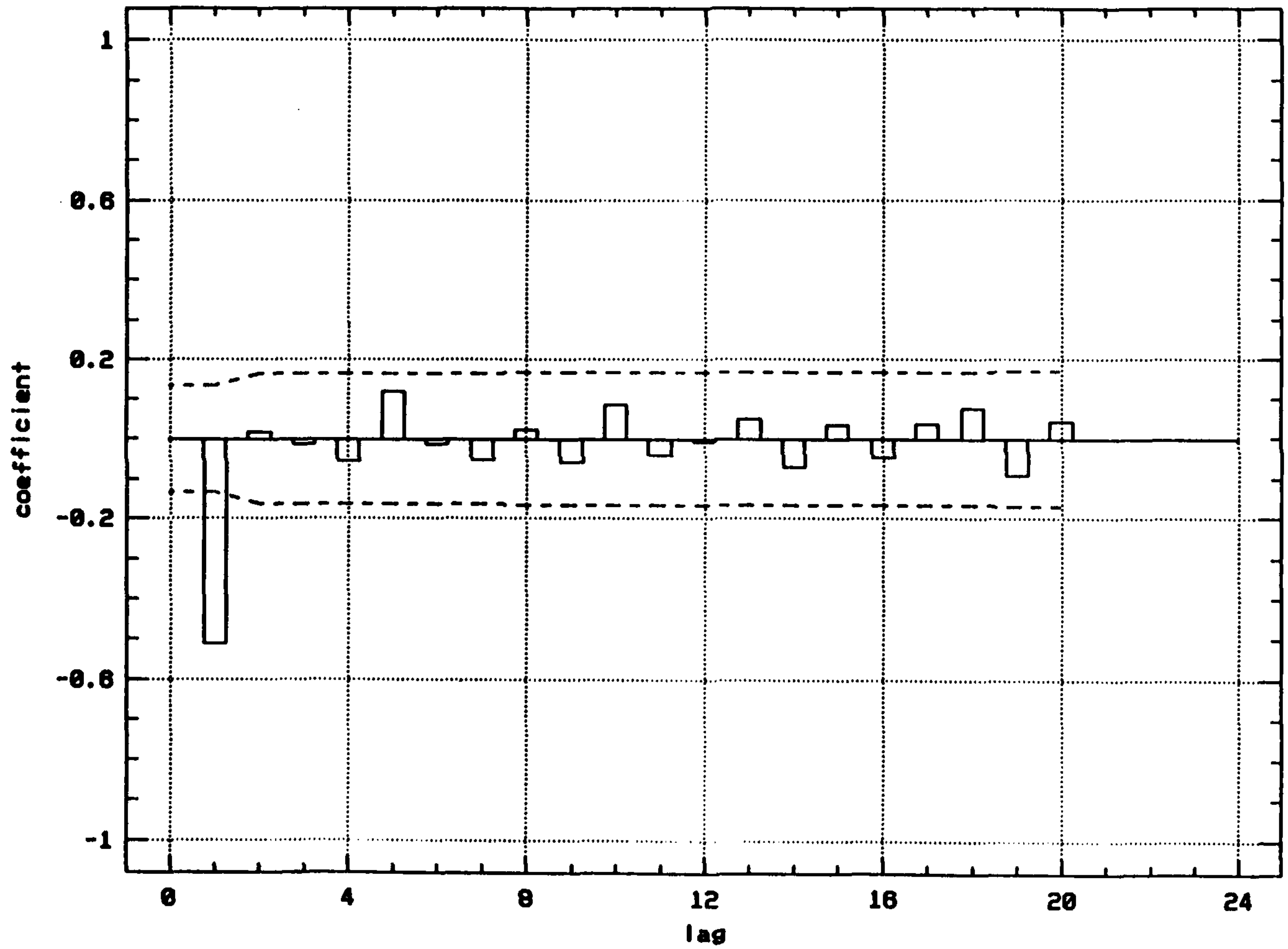


Figure 4.
Partial Autocorrelations
Asbestosis Claim Costs (Differenced at Lag 1)

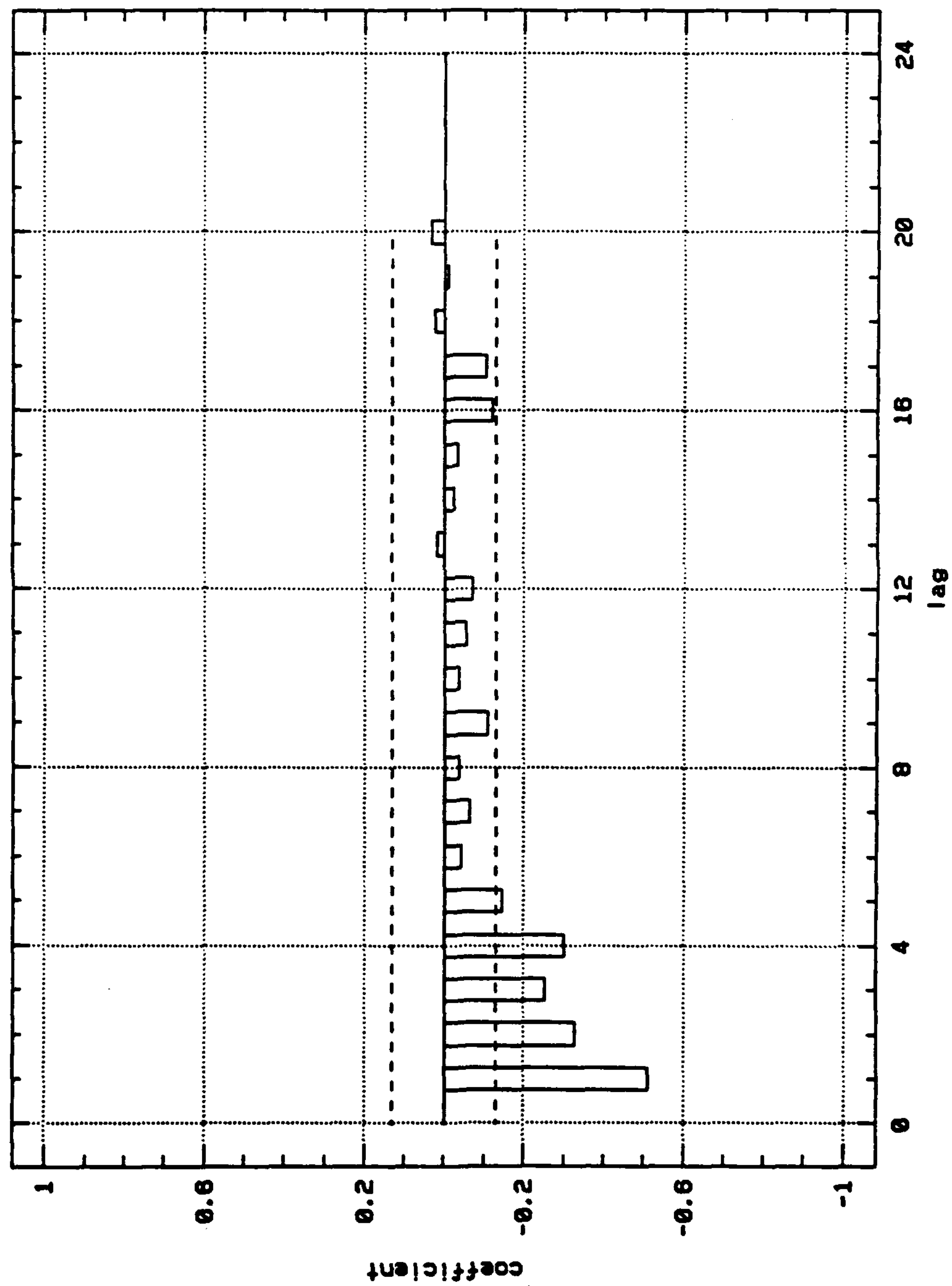
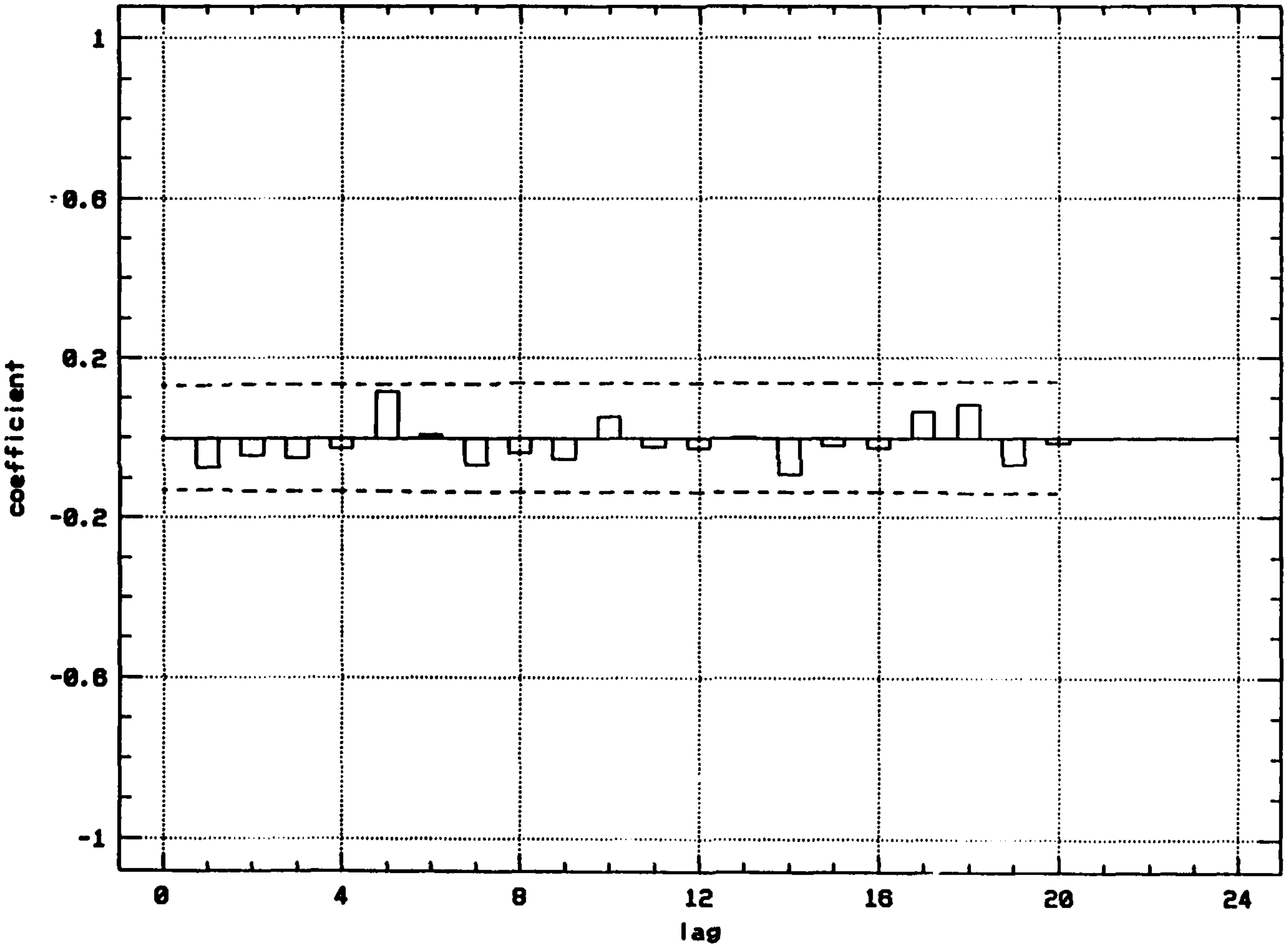


Figure 5.

Residual ACF
ARIMA(0,1,1) Fitted to Asbestosis Claim Costs



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 92

3. REPORT TYPE AND DATE COVERED
Final June 92 - August 92

4. TITLE AND SUBTITLE
Poisson Compounding of Dependent Random Variables:
A Stochastic Model for Total Claim Costs

5. FUNDING NUMBERS
Program Element:
Work Unit Number:
NAVFAC Reimbursable

6. AUTHOR(S)
Dr. John E. Angus

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No. 92-14

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

This research was undertaken on an ASEE Summer Faculty Research Fellowship.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In modeling total claims and compensation costs associated with hazardous material exposures, there are two sources of random variation. First, the claims are adjudicated at random points in time and secondly, the actual compensation awarded in each case is random. The Compound Poisson Process (CPP) is a useful model for describing total claim costs in the insurance industry when all types of claims are considered together. However, for specific claim types, the successive awards are likely to be correlated due to legal precedent, and various economic factors. The CPP is inaccurate in this situation. In this investigation, the CPP is modified to allow dependence among the compounding variables in an effort to more accurately model the claim awards. Specifically, it is assumed that the successive claim awards form an Autoregressive Integrated Moving Average (ARIMA) process. The justification and use of this model is illustrated using actual asbestosis claim cost data collected from Naval Shipyards.

14. SUBJECT TERMS
ARIMA models, hazardous materials exposure, asbestosis, central limit thereom for dependent random variables.

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102